Analyzing the Effect of Ensemble Size and Optimal Weighting of Machine Learning Models on the Accuracy of Monthly Rainfall Prediction

Abstract

This study investigates the effect of ensemble size on the predictive performance of machine learning models for monthly rainfall forecasting using the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) satellite precipitation dataset. A diverse set of machine learning algorithms was employed, including tree-based models (Random Forest, M5P, and M5Rules), regression-based models (Linear Regression and AdditiveRegression), kernel-based models (SMOreg and Gaussian Processes), instance-based models (IBk and LWL), and ensemble algorithms (Bagging, Random SubSpace, and Random Committee). Each model was independently trained, and their outputs were subsequently integrated using a weighted ensemble approach. For all possible combinations of 2 to 15 models, optimal weights were determined using an exhaustive Brute Force search, and the best ensemble was selected by maximizing the Nash–Sutcliffe Efficiency (NSE) coefficient. The results revealed a nonlinear relationship between ensemble size and predictive performance. The four-model ensemble consisting of SMOreg, AdditiveRegression, DecisionTable, and M5Rules achieved the best overall performance, with an NSE of 0.88 and an RMSE of 21.10 mm, providing the best balance between prediction accuracy, model stability, and computational efficiency. Adding models beyond this point did not improve predictive performance. Instead, increased error correlation, information redundancy, and performance saturation substantially increased computational time without meaningful accuracy gains. These findings indicate that structural diversity among base models is more important than simply increasing ensemble size. Although the quantitative results depend on dataset characteristics and climatic conditions, the proposed framework can be applied to similar regions after retraining with local data, providing a practical approach for developing efficient rainfall forecasting systems

.Keywords: *ML; Hydrological Modeling; Nash–Sutcliffe Efficiency (NSE); WEKA; CHIRPS.*

Introduction

Precipitation forecasting remains a cornerstone of hydrological and climatological sciences, playing a pivotal role in sustainable water resource management, agricultural planning, energy production, and natural disaster mitigation (Salahi et al., 2024). Given the increasing climatic volatility, accurate precipitation prediction has become essential for reducing the adverse impacts of extreme events, such as floods and droughts (Abdollahi et al., 2024; Fikileni & Wolski, 2020). However, the inherent limitations of traditional statistical models in capturing non-linear relationships and complex climatic behaviors have prompted a shift toward machine learning (ML) algorithms, which effectively uncover hidden patterns within complex datasets (Patel et al., 2023; Ramani et al., 2024). Despite advancements in modeling techniques, the sparsity of ground-based stations remains a critical challenge, particularly in arid and semi-arid regions. To address this, satellite-based datasets, such as CHIRPS, have emerged as robust sources for precipitation modeling, offering long-term temporal coverage (since 1981), high spatial resolution, and reliable integration of station and satellite data (Funk et al., 2015; Beck et al., 2017). Consequently, this study utilizes monthly CHIRPS data as the primary input for developing predictive ML models.

In recent decades, ML applications for forecasting hydrological variables, including runoff, streamflow, and groundwater levels, have grown significantly (Sharma et al., 2023). While individual models (e.g., SVM, MLP, RF, GP) often yield accurate results, their performance remains inconsistent when subjected to complex climatic variability (Zounemat Kermani, 2017). To mitigate this, ensemble learning approaches have been developed to enhance predictive accuracy and reduce the variance associated with single-model outputs (He et al., 2020; Mosavi et al., 2022). A fundamental challenge in ensemble design is determining the optimal number of constituent models. Increasing the ensemble size does not guarantee monotonic improvements in accuracy; frequently, systems reach a saturation point where adding further models yields diminishing returns or even degrades predictive performance (Kundu et al., 2023). Furthermore, the assignment of optimal weights remains a complex task, as improper weighting can severely undermine the ensemble's efficacy (Arsenault et al., 2015; Wang et al., 2023). Despite the

availability of various integration methods, such as simple averaging, Akaike-based criteria, and Bayesian model averaging, there is a lack of systematic research regarding the relationship between ensemble size and overall performance.

To fill this research gap, this study systematically investigates the influence of ensemble size on predictive accuracy, using the Nash-Sutcliffe Efficiency (NSE) index as the primary performance metric. By evaluating all possible model combinations (from 2 to 15) and employing an optimal weighting approach, we derive frequency distributions of NSE values to analyze performance trends and stability. This research aims to identify the optimal ensemble size and provide deeper insights into the transition mechanism from performance improvement to saturation, ultimately establishing a practical framework for the design of robust, ML-based hydrological forecasting systems.

Methodology

Data Acquisition

This study utilizes monthly precipitation data from the CHIRPS dataset for the Namak Lake basin, spanning January 2000 to December 2024. Developed by the Climate Hazards Group, CHIRPS provides precipitation data at a 0.05° spatial resolution. It integrates satellite infrared imagery, long-term climatologies, and ground-based station data. Due to its proven efficacy in data-scarce regions, CHIRPS serves as a reliable source for hydrological analysis (Funk et al., 2015).

Preprocessing and Data Splitting

Outlier detection was performed using the Interquartile Range (IQR) method (Tukey, 1977); statistical assessments confirmed the absence of significant outliers or discontinuities, thus eliminating the need for data normalization. The model inputs comprise monthly precipitation (mm) and month-index to capture seasonal effects. Data were partitioned into training (2000–2021) and testing (2022–2024) sets.

Validation and Machine Learning Models

The CHIRPS dataset was validated against ground observations from four meteorological stations (Aghcheh, Arteshabad, Araghaleh, and Ahmadabad) over a ten-year period (2007–2017). Initially, 16 machine learning models were evaluated; however, *Random Tree* was excluded due to poor performance ($R<0.5$). The remaining 15 models, representing diverse algorithmic approaches, including linear, distance-based, kernel-based (SMOreg, GP), neural networks (MLP, RBF), and tree-based models (M5P, M5Rules, Decision Table), were optimized using WEKA's *GridSearch* tool.

Ensemble Strategy

To enhance predictive accuracy and reduce stochastic errors, a weighted ensemble approach was implemented. We employed a *Brute-force* optimization algorithm to determine optimal weights (ranging from 0 to 1 with 0.01 increments) for all possible model combinations ($k = 2$ to 15). This approach identifies the combination that maximizes the Nash-Sutcliffe Efficiency (NSE) in the testing phase, enabling a systematic analysis of the "performance saturation" point in ensemble size. This methodology allows us to assess whether structural diversity, rather than merely increasing the number of models, is the primary driver of predictive performance.

Performance Evaluation Metrics

Model performance was evaluated using are RMSE, MAE, MSE, Bias and NSE.

Results

Accuracy Assessment of Satellite Precipitation Data

The accuracy of the CHIRPS dataset was validated against ground-based observations from four meteorological stations within the Namak Lake basin (2007–2017). Statistical analysis yielded a Nash-Sutcliffe Efficiency (NSE) of 70%, confirming the suitability of CHIRPS for hydrological modeling in this arid and semi-arid region. Despite a slight tendency toward underestimation (Bias = -0.09 mm), the dataset demonstrated sufficient reliability for the purposes of this study.

Performance of Individual Machine Learning Models

The performance of 15 selected ML algorithms was evaluated using the testing dataset (2022–2024). Models such as *M5Rules* (NSE: 80.81%) and *AdditiveRegression* (NSE: 75.2%) emerged as the most robust, characterized by negligible bias and lower error variance. In contrast, algorithms like *IBk* exhibited higher sensitivity to outliers, reflected in elevated Mean Squared Error (MSE) values. Our analysis highlights that while complex models (e.g., MLP) excel at identifying intricate, non-linear relationships, their performance is highly dependent on parameter tuning and data consistency. Tree-based models generally displayed superior stability in managing the non-stationary nature of monthly precipitation data.

Ensemble Performance and the Impact of Model Count

Our evaluation of all possible ensemble combinations (k = 2 to 15) revealed a non-linear relationship between ensemble size and predictive performance. The optimal ensemble configuration was identified as a four-model combination (*SMOreg, AdditiveRegression, DecisionTable, and M5Rules*), achieving a peak NSE of 88.76% and an RMSE of 21.10 mm.

The study establishes a clear "performance saturation" point: increasing the ensemble size beyond four models did not yield further accuracy gains; instead, it led to a gradual decline in performance and a sharp, exponential increase in computational costs, rising from approximately 1 minute for a dual-model ensemble to 320 minutes for the full 15-model combination. This suggests that the redundancy and noise introduced by excessive models may impede, rather than enhance, predictive accuracy.

Frequency Distribution Analysis of NSE

An analysis of the NSE frequency distribution further corroborates these findings. While smaller ensembles (2–4 models) exhibit high performance variability, ensembles containing 5–7 models demonstrate superior stability and a higher density of favorable outcomes within the 75–85% NSE range. Conversely, larger ensembles (k > 9) show a distinct "phase transition" toward saturation, where information redundancy and over-averaging diminish the potential for high-accuracy predictions. In conclusion, the four-model ensemble provides an optimal equilibrium between predictive quality, stability, and computational efficiency, demonstrating that structural diversity is significantly more critical than ensemble scale for hydrological forecasting.

Conclusions

This research investigates the impact of ensemble size on the predictive accuracy of machine learning precipitation forecasting, challenging the traditional assumption that larger ensembles inherently yield superior performance. Our analysis demonstrates a non-linear relationship between the number of constituent models (2–15) and forecasting efficacy. Findings reveal that while performance improves initially, a four-model ensemble (SMOreg, AdditiveRegression, DecisionTable, and M5Rules) serves as the performance apex (NSE: 88.76%). Beyond this threshold, the system encounters "saturation" and "information interference," leading to decreased accuracy and an exponential rise in computational overhead (from 1 to 320 minutes). These results underscore that strategic model selection, prioritizing optimization over numerical expansion, is a critical engineering necessity. Consequently, this study

advocates for the adoption of compact, high-efficiency ensembles in operational hydrology, offering a more stable and cost-effective paradigm for real-time environmental forecasting.

Author Contributions

All authors contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Data Availability Statement

Data are available on request from the authors.

Acknowledgements

The authors would like to thank the reviewers and editor for their critical comments that helped improve the paper. The authors gratefully acknowledge the support and facilities provided by the Irrigation and Development Group, Faculty of Agriculture, University of Tehran, Iran.

Ethical Considerations

The authors avoided data fabrication, falsification, plagiarism, and any form of scientific misconduct.

Conflict of Interest

The author declares no conflict of interest.

تحلیل اثر اندازه آنسامبل و وزن‌دهی بهینه مدل‌های یادگیری ماشین بر دقت پیش‌بینی بارش ماهانه بارش

چکیده

در این پژوهش، اثر تعداد اعضای انسامبل بر عملکرد مدل‌های یادگیری ماشین در پیش‌بینی بارش ماهانه با استفاده از داده‌های ماهواره‌ای CHIRPS بررسی شد. بدین منظور، طیفی از مدل‌های یادگیری ماشین شامل مدل‌های درختی(RandomForest) ، M5Pو(M5Rules ، رگرسیونی LinearRegression) و(AdditiveRegression ، کرنلی SMOreg) و(GaussianProcesses ، مبتنی بر فاصله IBk) و (LWL و ترکیبیBagging) ، RandomSubSpaceو (RandomCommitteeبه‌کار گرفته شدند و هر مدل به‌صورت مستقل آموزش داده شد. سپس خروجی مدل‌های پایه با استفاده از روش انسامبل وزن‌دار ترکیب شد. برای تمامی ترکیب‌های ممکن شامل ۲ تا ۱۵ مدل، وزن‌های بهینه با روش جستجوی فراگیر (Brute Force) تعیین و بهترین ترکیب بر اساس بیشینه‌سازی شاخص NSE انتخاب شد. هدف پژوهش، تعیین تعداد بهینه مدل‌های انسامبل و بررسی رابطه میان اندازه انسامبل، دقت پیش‌بینی و هزینه محاسباتی بود. نتایج نشان داد که رابطه میان تعداد مدل‌ها و دقت پیش‌بینی غیرخطی است. انسامبل چهارمدلی شاملSMOreg ، AdditiveRegression، DecisionTableو M5Rules با NSE برابر ۰/۸۸ و RMSE برابر ۲۱/۱۰، بهترین تعادل را میان دقت، پایداری عملکرد و هزینه محاسباتی ایجاد کرد. افزایش تعداد مدل‌ها فراتر از این نقطه، به دلیل همبستگی خطاها، هم‌پوشانی اطلاعات و اشباع عملکرد، بهبود معناداری در دقت ایجاد نکرد و در مقابل، زمان محاسباتی را افزایش داد. یافته‌ها نشان می‌دهد تنوع ساختاری مدل‌های پایه اهمیت بیشتری از افزایش تعداد آن‌ها دارد. چارچوب پیشنهادی می‌تواند پس از آموزش مجدد با داده‌های منطقه هدف، برای طراحی سامانه‌های پیش‌بینی بارش در مناطق با شرایط اقلیمی مشابه به‌کار رود.

کلید واژه: *یادگیری ماشین، انسامبل، مدلسازی هیدرولوژی، شاخص نش–ساتکلیف، WEKA ، CHIRPS*

مقدمه

پیش‌بینی بارش به‌عنوان یکی از مهم‌ترین چالش‌های علوم هیدرولوژی و اقلیم‌شناسی، نقش بنیادینی در مدیریت پایدار منابع آب، کشاورزی، تولید انرژی و کاهش خطرات طبیعی دارد (Salahi et al., 2024). با افزایش نوسانات اقلیمی و تغییر در الگوهای بارش،

دستیابی به پیش‌بینی‌های دقیق برای کاهش اثرات مخرب پدیده‌هایی همچون سیلاب و خشکسالی ضروری شده است (Abdollahi et al., 2024; Fikileni & Wolski, 2020). کاهش دقت مدل‌های آماری سنتی در برابر روابط غیرخطی و رفتارهای پیچیده اقلیمی موجب شده است که پژوهش‌های اخیر به سمت استفاده از الگوریتم‌های یادگیری ماشین[1]سوق پیدا کنند؛ الگوریتم‌هایی که با قابلیت شناسایی الگوهای پنهان در داده‌ها، امکان مدل‌سازی دقیق‌تر متغیرهای هیدرولوژیکی را فراهم می‌سازند (Patel et al., 2023; Ramani et al., 2024).

با وجود پیشرفت در روش‌های پیش‌بینی، کیفیت و تراکم پایین ایستگاه‌های زمینی همچنان یکی از چالش‌های اصلی در مناطق خشک و نیمه‌خشک است. در این میان، داده‌های ماهواره‌ای نظیر CHIRPS[2] به دلیل پوشش زمانی بلندمدت (۱۹۸۱ تاکنون)، دقت مکانی حدود ۵ کیلومتر و تلفیق داده‌های ایستگاهی و ماهواره‌ای، به‌عنوان یکی از معتبرترین منابع داده بارش برای مطالعات هیدرولوژیکی و اقلیمی شناخته می‌شوند (Funk et al., 2015; Tyralis et al., 2021). مطالعات اخیر نیز کارایی مناسب این محصول را در برآورد بارش، پایش خشکسالی و مدل‌سازی هیدرولوژیکی در مناطق مختلف جهان تأیید کرده‌اند و نشان داده‌اند که CHIRPS در بسیاری از مناطق دارای کمبود ایستگاه‌های زمینی، عملکرد قابل قبولی در بازنمایی تغییرات زمانی و مکانی بارش دارد (Ayugi et al., 2022) ویژگی‌هایی نظیر دسترسی آزاد، به‌روزرسانی مستمر، سازگاری با سامانه‌های اطلاعات جغرافیایی و قابلیت استفاده در مدل‌های یادگیری ماشین، CHIRPS را به یکی از پرکاربردترین محصولات ماهواره‌ای در مطالعات پیش‌بینی بارش تبدیل کرده است (Ayugi et al., 2022). ازاین‌رو، در پژوهش حاضر، داده‌های ماهانه CHIRPS به‌عنوان ورودی اصلی مدل‌های یادگیری ماشین برای پیش‌بینی بارش ماهانه مورد استفاده قرار گرفت..

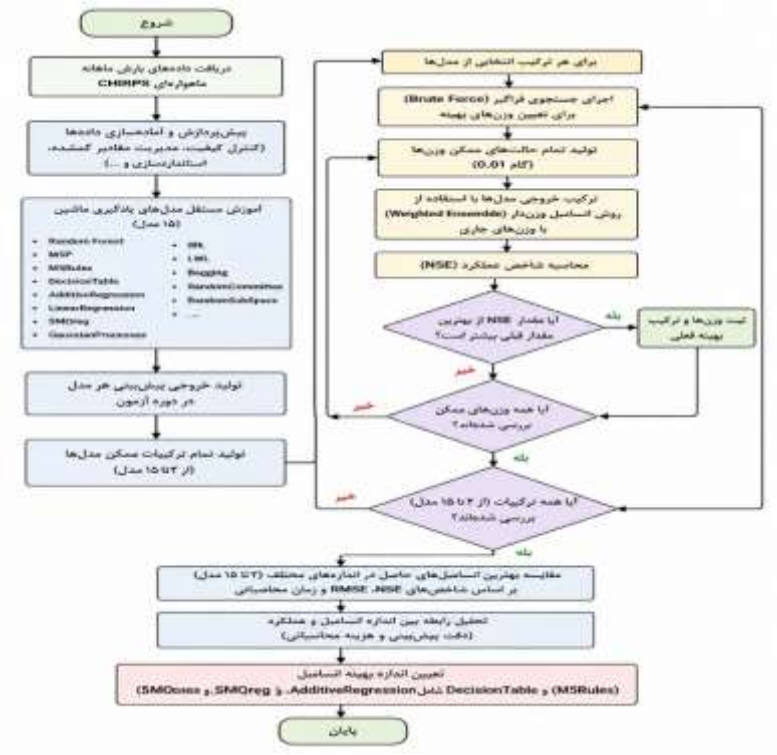
در دهه‌های اخیر، کاربرد مدل‌های یادگیری ماشین برای پیش‌بینی متغیرهای هیدرولوژیکی همانند بارش، رواناب، دبی رودخانه و سطح آب زیرزمینی رشد چشمگیری یافته است (دلیر گبرآباد و همکاران ۱۴۰۵) مطالعات نشان داده‌اند که اگرچه مدل‌های مختلف مانند SVM[3]، MLP[4]، RF[5] یا GP[6] می‌توانند نتایج دقیقی ارائه کنند، اما عملکرد تک‌مدل‌ها در مواجهه با پیچیدگی‌های اقلیمی و تغییرپذیری داده‌ها همواره یکسان و پایدار نیست (Zounemat-Kermani, 2017.) به همین دلیل، رویکردهای مبتنی بر ترکیب مدل‌ها (Ensemble Learning) به عنوان یک راهکار برای افزایش دقت و کاهش واریانس مدل‌های منفرد توسعه یافته‌اند (Tyralis et al., 2021). این رویکردها با بهره‌گیری از تنوع الگوریتمی، خطاهای مدل‌های مختلف را تا حد قابل توجهی جبران کرده و دقت پیش‌بینی را ارتقا می‌دهند (He et al., 2020).

در مدل‌های انسامبل، یکی از تعیین‌کننده‌ترین مؤلفه‌ها تعداد مدل‌های ترکیب‌شده است؛ زیرا افزایش تعداد مدل‌ها همیشه منجر به بهبود دقت نمی‌شود و در بسیاری از موارد، سیستم به نقطه اشباع می‌رسد که در آن افزودن مدل‌های بیشتر نه تنها کمکی به افزایش دقت نمی‌کند، بلکه ممکن است اثر منفی نیز بر خروجی نهایی داشته باشد (Kundu et al., 2023). از سوی دیگر، تخصیص وزن‌های بهینه به مدل‌ها یکی از چالش‌های اصلی این روش است، چراکه وزن‌دهی نادرست می‌تواند عملکرد نهایی را حتی از مدل‌های منفرد نیز ضعیف‌تر کند (Arsenault et al., 2015; Wang et al., 2023). اگرچه روش‌های مختلفی مانند میانگین‌گیری ساده، روش‌های مبتنی بر معیار آکائیک و میانگین‌گیری بیزین برای ترکیب مدل‌ها معرفی شده‌اند، اما تمرکز اغلب پژوهش‌های پیشین بر توسعه یا مقایسه روش‌های وزن‌دهی بوده و تأثیر تعداد اعضای انسامبل بر عملکرد پیش‌بینی، به‌صورت نظام‌مند و جامع مورد بررسی قرار نگرفته است. به‌ویژه، در مطالعات پیشین مشخص نشده است که افزایش تعداد مدل‌های مشارکت‌کننده تا چه نقطه‌ای موجب بهبود دقت می‌شود و

[1] Machine Learning
[2] CHIRPS: Climate Hazards Group InfraRed Precipitation with Station data
[3] Support Vector Machine
[4] Multi-Layer Perceptron
[5] Random Forest
[6] Gaussian Processes

از چه مرحله‌ای به بعد، به دلیل همپوشانی اطلاعات، افزایش پیچیدگی و کاهش تنوع مؤثر مدل‌ها، عملکرد انسامبل وارد ناحیه اشباع یا حتی افت دقت می‌شود. افزون بر این، در زمینه پیش‌بینی بارش ماهانه نیز مطالعات اندکی به ارزیابی همزمان اثر تعداد مدل‌ها، وزن‌دهی بهینه و هزینه محاسباتی پرداخته‌اند. ازاین‌رو، شکاف اصلی پژوهش حاضر، نبود یک ارزیابی جامع از ارتباط میان اندازه انسامبل، وزن‌دهی بهینه و دقت پیش‌بینی در چارچوب مدل‌های یادگیری ماشین است. بر همین اساس، پژوهش حاضر با بررسی تمامی ترکیب‌های ممکن از مدل‌های منتخب و تعیین وزن‌های بهینه برای هر ترکیب، درصدد شناسایی تعداد بهینه مدل‌های انسامبل و تبیین رابطه میان اندازه انسامبل، دقت پیش‌بینی و هزینه محاسباتی در پیش‌بینی بارش ماهانه است.

در همین راستا، پژوهش حاضر با هدف تحلیل نظام‌مند اثر تعداد مدل‌ها بر عملکرد انسامبل‌های یادگیری ماشین، از شاخص NSE به‌عنوان معیار اصلی ارزیابی استفاده کرده و به این پرسش اساسی پاسخ می‌دهد که «افزایش تعداد مدل‌های انسامبل تا چه حد موجب بهبود دقت پیش‌بینی می‌شود و از چه نقطه‌ای به بعد، عملکرد مدل وارد ناحیه اشباع یا افت کارایی می‌شود؟». برخلاف اغلب مطالعات پیشین که تنها تعداد محدودی از ترکیب‌های مدل یا روش‌های مختلف وزن‌دهی را مورد ارزیابی قرار داده‌اند، در این پژوهش تمامی ترکیب‌های ممکن مدل‌های منتخب، از ترکیب‌های دومدلی تا ترکیب کامل ۱۵مدلی، به‌صورت جامع بررسی شده‌اند. همچنین، برای هر ترکیب، وزن‌های بهینه مدل‌ها با استفاده از رویکرد جستجوی فراگیر (Brute Force) تعیین شده و توزیع فراوانی مقادیر NSE استخراج گردیده است تا رفتار تغییرات دقت، پایداری پیش‌بینی و روند گذار از بهبود عملکرد به اشباع به‌صورت کمی تحلیل شود. نوآوری اصلی این پژوهش، ارائه یک چارچوب جامع برای ارزیابی همزمان اثر اندازه انسامبل، وزن‌دهی بهینه و هزینه محاسباتی بر عملکرد پیش‌بینی بارش است که علاوه بر شناسایی تعداد بهینه مدل‌ها، مبنایی علمی برای طراحی انسامبل‌های کارآمد و جلوگیری از افزایش غیرضروری پیچیدگی محاسباتی فراهم می‌سازد.



شکل ۱. روندنمای پژوهش

پیشینه پژوهش

در دهه‌های اخیر، روش‌های یادگیری ماشین و به‌ویژه سامانه‌های انسامبل به دلیل توانایی در افزایش دقت، پایداری و قابلیت تعمیم، به‌طور گسترده در پیش‌بینی‌های هیدرولوژیکی استفاده شده‌اند. برخلاف مدل‌های منفرد که تحت تأثیر محدودیت داده، عدم‌قطعیت پارامترها، نویز و رخدادهای حدی قرار می‌گیرند، انسامبل‌ها با ترکیب مدل‌های دارای ساختار، داده آموزشی یا شرایط اولیه متفاوت می‌توانند بخشی از خطاها را کاهش دهند. در این میان، اندازه انسامبل و تنوع اعضا از مهم‌ترین عوامل تعیین‌کننده عملکرد هستند. اگرچه افزایش تعداد اعضا معمولاً موجب کاهش واریانس و افزایش پایداری می‌شود، این اثر خطی نیست و پس از رسیدن به نقطه‌ای مشخص، عملکرد اشباع می‌شود. در مقابل، ترکیب مدل‌های دارای الگوهای خطای متفاوت می‌تواند اطلاعات مکمل ایجاد کند، در حالی‌که مدل‌های مشابه عمدتاً موجب افزونگی اطلاعات و افزایش هزینه محاسباتی می‌شوند.

مطالعات نشان داده‌اند که انسامبل‌ها می‌توانند بایاس و واریانس را کاهش داده و پایداری پیش‌بینی را در شرایط متغیر هیدرولوژیکی افزایش دهند Snieder et al. (2020) .نشان دادند که روش‌های انسامبل نسبت به مدل‌های ساده ANN[1] خطاهای دامنه و زمان‌بندی را کاهش می‌دهند. همچنین، Hartanto (2019)گزارش کرد که استفاده از اعضای انسامبل در شبیه‌سازی احتمالاتی دبی، توانایی تشخیص عبور از آستانه‌های بحرانی را بهبود می‌دهد. از دیدگاه نظری، شواهد نشان می‌دهند که برتری انسامبل‌ها بیشتر به تنوع ساختاری وابسته است تا صرفاً تعداد اعضا Sharma et al. (2019) .نشان دادند که هر مدل می‌تواند اطلاعات متفاوتی به فرآیند پیش‌بینی اضافه کند، در حالی‌که Darbandsari et al. (2019) گزارش کردند حتی مدل‌های ضعیف‌تر نیز در صورت ارائه اطلاعات مکمل می‌توانند عملکرد انسامبل را بهبود دهند. مطالعات مختلف نیز وجود رابطه‌ای غیرخطی میان اندازه انسامبل و عملکرد را تأیید کرده‌اند Na et al. (2019) .شش عضو را برای دستیابی به عملکرد مناسب کافی دانستند، در حالی‌که Ho et al. (2021) نشان دادند که بخش عمده بهبود در شبیه‌سازی رسوب تا حدود ۲۰ عضو حاصل شده و افزایش تعداد اعضا فراتر از آن، بهبود محدودی ایجاد می‌کند Palma et al. (2025) .نیز عملکرد نزدیک به بهینه را با حدود ۲۰ عضو گزارش کردند. با این حال، در مسائل پیچیده‌تر، انسامبل‌های بزرگ ممکن است مزیت بیشتری داشته باشند Rasmussen et al. (2015) و DeChant et al. (2011) نیز بر وجود مبادله میان دقت و هزینه پردازشی در افزایش تعداد اعضا تأکید کردند.

انتخاب و وزن‌دهی اعضا نیز نقش مهمی در عملکرد انسامبل دارد Figueiredo et al. (2017) .نشان دادند که نحوه وزن‌دهی می‌تواند به‌اندازه تعداد اعضا بر عملکرد اثرگذار باشد. روش BMA[2] نیز با تعیین وزن مدل‌ها بر اساس عملکرد آن‌ها، علاوه بر بهبود پیش‌بینی قطعی، قابلیت نمایش عدم‌قطعیت را افزایش می‌دهد (Baran et al., 2018). مطالعات جدیدتر نشان داده‌اند که وزن‌دهی می‌تواند به‌عنوان یک مسئله یادگیری نیز در نظر گرفته شود (Frame et al., 2025; Peng et al., 2025) ، اما وزن‌دهی زمانی مؤثر است که اعضا اطلاعات مکمل ارائه دهند.

شواهد تجربی در پیش‌بینی هیدرولوژیکی نیز برتری انسامبل‌ها را نسبت به مدل‌های منفرد تأیید کرده‌اند Zhai et al. (2022) .دقت بالاتر و خطای نسبی کمتر انسامبل‌ها را در پیش‌بینی دبی روزانه گزارش کردند Sheikh et al. (2024) .نیز در یک مطالعه مروری نتیجه گرفتند که مدل‌های هیبریدی و انسامبل در بسیاری از حوضه‌ها عملکرد بهتری از مدل‌های منفرد دارند. در حوزه یادگیری عمیق، Armstrong et al. (2025)و Papacharalampous et al. (2022) نشان دادند که مدل‌های LSTM[3] در ترکیب با سایر مدل‌ها می‌توانند عملکرد بهتری داشته باشند. همچنین، Valdez et al. (2021)نشان داد که افزایش صرف تعداد اعضا الزاماً بهبود عملکرد را تضمین نمی‌کند، در حالی‌که Sharma et al. (2021) و Bellier et al. (2021) بر اهمیت تنوع ناشی از منابع مختلف داده‌های هواشناسی تأکید

[1] Artificial Neural Network
[2] Bayesian Model Averaging
[3] Long Short-Term Memory

کردند Cloke et al. (2017) و McMillan et al. (2013) نیز نقش تنوع اعضا را در آشکارسازی رخدادهای حدی و حفظ پایداری پیش‌بینی برجسته کردند. نتایج مطالعات درباره اندازه بهینه انسامبل نیز متفاوت است Ho et al. (2021) و Sikorska-Senoner et al. (2021)وجود افزونگی اطلاعاتی در انسامبل‌های بزرگ را نشان دادند؛ به‌طوری‌که تعداد محدودی عضو می‌تواند عملکردی مشابه انسامبل‌های بسیار بزرگ ایجاد کند. در مقابل، Necker et al. (2024)نشان دادند که برای نمونه‌برداری مناسب از ریسک رخدادهای حدی جوی، ممکن است بیش از ۲۰۰ عضو مورد نیاز باشد. همچنین، Bouallegue et al. (2020) نشان دادند که زیرمجموعه‌های کوچک اما مکمل می‌توانند عملکردی مشابه انسامبل کامل داشته باشند Seo et al. (2025) و Shi et al. (2025) نیز بر مزیت انسامبل‌های کوچک و هدفمند از نظر دقت و هزینه محاسباتی تأکید کردند. بنابراین، اندازه بهینه انسامبل به نوع مسئله، افق پیش‌بینی و هدف مدل‌سازی وابسته است.

با وجود گسترش کاربرد انسامبل‌ها در پیش‌بینی هیدرولوژیکی، بخش عمده مطالعات بر مقایسه مدل‌های منفرد و انسامبل، توسعه روش‌های وزن‌دهی یا بهبود ساختار مدل‌های ترکیبی متمرکز بوده‌اند و اثر نظام‌مند تعداد اعضای انسامبل بر پیش‌بینی بارش مبتنی بر داده‌های ماهواره‌ای کمتر بررسی شده است. همچنین، بسیاری از مطالعات دقت را بدون بررسی همزمان رابطه میان اندازه انسامبل، تنوع مدل‌ها و هزینه محاسباتی ارزیابی کرده‌اند. از این رو، پژوهش حاضر با استفاده از داده‌های ماهواره‌ای CHIRPS و ارزیابی پیکربندی‌های مختلف انسامبل بر اساس معیارهای NSE و RMSE، به دنبال تعیین نقطه بهینه میان دقت، پایداری و هزینه محاسباتی است. نوآوری اصلی پژوهش، تحلیل مستقیم اثر تعداد مدل‌های یادگیری ماشین بر عملکرد انسامبل و شناسایی مرز بهینه پارتو در طراحی سامانه پیش‌بینی بارش است؛ به‌گونه‌ای که مشخص می‌شود افزایش تعداد مدل‌ها الزاماً موجب بهبود عملکرد نمی‌شود و تنوع ساختاری اعضا می‌تواند اهمیت بیشتری از افزایش صرف تعداد آن‌ها داشته باشد.

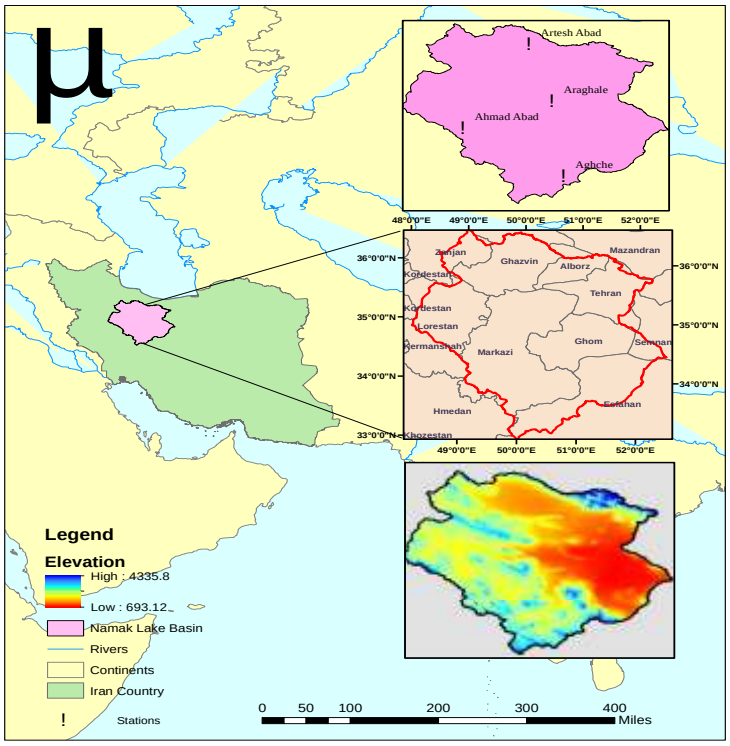
روش شناسی پژوهش

داده های مورد استفاده

در این پژوهش، از داده‌های بارش ماهانه پایگاه CHIRPS برای حوضه آبریز دریاچه نمک در بازه زمانی ژانویه ۲۰۰۰ تا دسامبر ۲۰۲۴ استفاده شد (شکل ۱). مجموعه داده CHIRPS که توسط گروه مخاطرات اقلیمی[1] توسعه یافته است، داده‌های بارش را با تفکیک مکانی ۰/۰۵ درجه از سال ۱۹۸۱ تاکنون ارائه می‌دهد. این پایگاه داده از تلفیق تصاویر مادون‌قرمز ماهواره‌ای، اقلیم‌نگاشت‌های بلندمدت و داده‌های ایستگاه‌های باران‌سنجی زمینی بهره می‌برد. فرآیند تولید این داده‌ها شامل برآورد بارش بر اساس دمای ابر (سنجش از دور)، کالیبراسیون با داده‌های ایستگاهی و تصحیح بایاس به روش‌های آماری است. به دلیل عملکرد مطلوب این داده‌ها در مناطق دارای شبکه ایستگاهی محدود و دقت بالا در مناطق خشک و نیمه‌خشک، CHIRPS به عنوان یک منبع داده‌ای معتبر برای تحلیل‌های هیدرولوژیک حوضه آبریز دریاچه نمک انتخاب شد (Funk . et al., 2015) در ادامه شکل ۲ موقعیت حوضه مطالعاتی را نشان می‌دهد.

[1] Climate Hazards Group

aerA ydutS



شکل ۲. حوضه دریاچه نمک

پیش پردازش و تقسیم بندی داده ها

پیش از مرحله مدل‌سازی، داده‌های بارش تحت عملیات پیش‌پردازش قرار گرفتند. برای شناسایی داده‌های پرت، از روش دامنه بین چارکی[1]استفاده شد. در این روش، پس از محاسبه چارک‌های اول (Q1) و سوم (Q3)، داده‌هایی که خارج از بازه [a,b] قرار می‌گرفتند، به عنوان داده‌های پرت شناسایی شدند (Turkey, 1977). بررسی‌های آماری نشان داد که سری زمانی فاقد داده‌های پرت معنادار است. همچنین، ارزیابی روند و پیوستگی سری زمانی عدم وجود گسست غیرمنطقی را تأیید کرد؛ از این رو، با توجه به یکنواختی مقیاس داده‌ها، نیازی به نرمال‌سازی مشاهده نشد.

رابطه ۱) $IQR = Q3 - Q1$

رابطه ۲) کران بالا $(a) = Q1 - 1.5 \times IQR$

رابطه ۳) کران پایین $(b) = Q3 + 1.5 \times IQR$

ورودی‌های مدل شامل میانگین بارش ماهانه (میلی‌متر) و شماره ماه (به‌منظور لحاظ کردن اثرات فصلی) بود. در این مطالعه از تأخیر زمانی[2]استفاده نشد تا تمرکز اصلی بر رابطه میان بارش ماهانه و نوسانات فصلی باقی بماند. داده‌ها به دو بخش آموزش (ژانویه ۲۰۰۰ تا دسامبر ۲۰۲۱) و آزمون (ژانویه ۲۰۲۲ تا دسامبر ۲۰۲۴) تقسیم شدند.

ارزیابی دقت داده‌های بارش CHIRPS

[1] Interquartile Range – IQR
[2] Time Lag

به‌منظور صحت‌سنجی داده‌های ماهواره‌ای، از داده‌های مشاهده‌ای چهار ایستگاه زمینی (آغچه، ارتش‌آباد، آراقلعه و احمدآباد) در دوره آماری ۱۰ ساله (۲۰۰۷ تا ۲۰۱۷) استفاده شد (مشخصات در جدول ۱). پس از استخراج مقادیر متناظر با مختصات ایستگاه‌ها، شاخص‌های آماری برای هر ایستگاه به‌صورت مجزا محاسبه و سپس میانگین آن‌ها برای تعیین عملکرد کلی داده‌های CHIRPS در سطح حوضه ارائه گردید (جدول ۲). همچنین موقعیت ایستگاه‌ها در شکل ۲ نشان‌داده شده است.

جدول ۱. مختصات ایستگاه‌های بررسی شده

نام ایستگاه	استان	عرض جغرافیایی	طول جغرافیایی	ارتفاع (متر)
آغچه	اصفهان	۳۳-۰۴-۰۲	۵۰-۲۳-۲۵	۲۵۰۰
ارتش آباد	قزوین	۳۵-۳۹-۳۱	۴۹-۲۵-۴۷	۱۷۵۰
آراقلعه	مرکزی	۳۵-۰۶-۰۵	۴۹-۴۳-۱۹	۱۶۴۱
احمدآباد	مرکزی	۳۴-۵۹-۴۲	۵۰-۲۶-۱۸	۱۰۴

در ابتدای فرآیند، ۱۶ مدل یادگیری ماشین مورد ارزیابی قرار گرفتند. پس از تحلیل اولیه، مدل Random Tree به دلیل عملکرد ضعیف (ضریب همبستگی کمتر از ۰/۵ و RMSE بالا) حذف و ۱۵ مدل برتر برای تحلیل نهایی انتخاب شدند. الگوریتم‌های انتخاب شده طیف متنوعی از رویکردهای یادگیری ماشین را شامل می‌شوند:

- مدل‌های خطی: Linear Regression
- مدل‌های مبتنی بر فاصله: IBk
- مدل‌های مبتنی بر هسته (Kernel-based): SMOreg و Gaussian Processes (GP)
- مدل‌های شبکه عصبی: MLP و RBF
- مدل‌های درختی و قاعده‌محور: M5P، M5Rules و Decision Table
- مدل‌های ترکیبی: Random Forest، Random SubSpace، Random Committee، Additive Regression و DecisionTable

انتخاب این مجموعه متنوع با هدف ارزیابی قابلیت مدل‌ها در بازنمایی رفتارهای غیرخطی و نوسانی بارش و همچنین بررسی اثر انواع مدل‌ها بر روی یکدیگر به دلیل قابلیت کاهش واریانس و مدیریت روابط پیچیده، عملکردی فراتر از مدل‌های پایه دارند. در نهایت، پارامترهای هر مدل با استفاده از ابزار GridSearch در نرم‌افزار WEKA[1] به صورت بهینه تنظیم شد. جزئیات مدل‌های مورد استفاده در جدول ۲ ارائه شده است.

جدول ۲. توضیحات مدل‌های یادگیری ماشین مورد استفاده در پژوهش

مدل	مزایا	معایب	مرجع

[1] Waikato Environment for Knowledge Analysis

GaussianProcesses	توانایی مدل‌سازی روابط غیرخطی پیچیده؛ ارائه عدم‌قطعیت پیش‌بینی؛ عملکرد قوی در داده‌های کوچک	هزینه محاسباتی بالا، حساس به انتخاب کرنل	(Schulz et al., 2018)
RandomForest	مقاوم به نویز؛ کاهش بیش‌برازش نسبت به درخت منفرد؛ عملکرد قوی در داده‌های غیرخطی	کاهش تفسیرپذیری؛ نیاز به تنظیم تعداد درخت‌ها و پارامترهای دیگر	(Biau, G., & Scornet, E. 2016)
MLPRegressor	قدرت بالا در مدل‌سازی روابط غیرخطی پیچیده؛ مناسب سری‌های زمانی	نیاز به تنظیم زیاد پارامترها؛ حساس به داده کم	(Goodfellow, I., Bengio, Y., & Courville, A. 2016)
LinearRegression	ساده، سریع، قابل تفسیر؛ مبنای مقایسه	ناتوان در مدل‌سازی روابط غیرخطی	(James, G. et al., 2021)
SMOreg	عملکرد پایدار در داده‌های نویزی؛ کنترل بیش‌برازش	انتخاب پارامتر کرنل پیچیده	(Cervantes, J et al 2020)
IBk	ساده؛ بدون فرض توزیع؛ مناسب داده‌های محلی	حساس به مقیاس داده؛ عملکرد ضعیف در داده پرتعداد	(Taunk, K et al 2019)
LWL	مدل‌سازی رفتار موضعی؛ انعطاف‌پذیر	هزینه محاسباتی بالا در پیش‌بینی؛ حساس به پارامتر وزن	(Alpaydin, E. 2020)
AdditiveRegression	کاهش بایاس؛ بهبود مدل پایه	خطر بیش‌برازش در تکرار زیاد	(Chen, T., & Guestrin, C. 2016)
Bagging	کاهش واریانس؛ پایداری بالا	کاهش تفسیرپذیری	(Sagi, O., & Rokach, L. 2018)
RandomCommittee	بهبود پایداری مدل پایه؛ ساده در اجرا	وابسته به کیفیت مدل پایه	(Dong, X. et al 2020)
RandomSubSpace	کاهش همبستگی بین مدل‌ها؛ مناسب داده‌های با ویژگی زیاد	احتمال حذف ویژگی‌های مهم	(Kuncheva, L. I. 2014)
DecisionTable	تفسیرپذیری بالا؛ سادگی	عملکرد ضعیف در روابط پیچیده	(Fürnkranz, J et al 2012)
M5Rules	ترکیب درخت تصمیم و رگرسیون خطی؛ تفسیرپذیر و دقیق	حساس به داده‌های نویزی	(Holmes, G et al 1999)
M5P	مناسب روابط غیرخطی؛ ساختار قابل تفسیر	امکان بیش‌برازش در داده کم	(Wang, Y., & Witten, I. H. 1997)
RBFRegressor	همگرایی سریع‌تر از MLP؛ مناسب روابط موضعی	حساس به انتخاب مراکز و پارامتر	(Wu, Y et al 2012)

روش ترکیب مدل‌ها (Ensemble Strategy)

به‌منظور دستیابی به بالاترین دقت در پیش‌بینی بارش و کاهش خطاهای تصادفی مدل‌های منفرد، از یک رویکرد ترکیب وزن‌دار (Weighted Ensemble) استفاده شد. در این روش، ابتدا هر یک از ۱۵ مدل یادگیری ماشین به‌صورت مستقل آموزش داده شده و مقادیر پیش‌بینی‌شده آن‌ها برای دوره آزمون استخراج گردید. سپس خروجی مدل‌ها به‌جای استفاده مستقیم، با یکدیگر ادغام شدند؛ به‌گونه‌ای که هر مدل سهم متفاوتی در پیش‌بینی نهایی داشته باشد .

به‌طور مشخص، اگر خروجی مدل i ام برابر با P_i و وزن متناظر آن برابر با W_i باشد، مقدار نهایی پیش‌بینی انسامبل از رابطه زیر محاسبه شد:

$$P_{\text{ensemble}} = \sum_{i=1}^{k} W_i P_i$$ رابطه ۴)

که در آن k تعداد مدل‌های حاضر در ترکیب است. به‌منظور حفظ مقیاس پیش‌بینی‌ها، مجموع وزن‌های اختصاص‌یافته به مدل‌های هر ترکیب برابر با یک در نظر گرفته شد؛ به عبارت دیگر:

رابطه ۵)
$$\sum_{i=1}^{k} W_i = 1$$

بنابراین، مدل‌هایی که عملکرد بهتری داشتند، در صورت انتخاب توسط فرآیند بهینه‌سازی، وزن بیشتری دریافت کرده و تأثیر بیشتری بر خروجی نهایی انسامبل داشتند، در حالی که سهم مدل‌های ضعیف‌تر به‌صورت خودکار کاهش یافت.. برای تعیین بهترین وزن‌ها، از روش جستجوی فراگیر (Brute-force Search) استفاده شد. در این روش، برای هر ترکیب از مدل‌ها، تمامی حالت‌های ممکن وزن‌ها با گام ۰/۰۱ تولید شدند. به عبارت دیگر، وزن هر مدل مقادیری بین ۰ تا ۱ با افزایش‌های ۰/۰۱ اختیار می‌کرد و تنها حالت‌هایی پذیرفته شدند که مجموع وزن‌ها برابر یک باشد. برای هر مجموعه وزن، ابتدا خروجی انسامبل با استفاده از میانگین وزنی محاسبه شد، سپس شاخص کارایی نش–ساتکلیف (NSE) در دوره آزمون محاسبه گردید. در نهایت، مجموعه وزن‌هایی که بیشترین مقدار NSE را ایجاد می‌کرد، به‌عنوان وزن‌های بهینه همان ترکیب انتخاب شد. این فرآیند برای تمامی ترکیب‌های ممکن مدل‌ها، از ترکیب‌های دومدلی تا ترکیب کامل ۱۵مدلی $\binom{15}{k}$، تکرار شد. بدین ترتیب، برای هر تعداد مدل، بهترین ترکیب و بهترین وزن‌ها به‌طور مستقل تعیین گردید و امکان مقایسه منصفانه عملکرد انسامبل‌ها با اندازه‌های مختلف فراهم شد. این رویکرد علاوه بر استخراج وزن‌های بهینه، امکان بررسی این پرسش اساسی را نیز فراهم ساخت که آیا افزایش تعداد مدل‌های انسامبل همواره موجب بهبود عملکرد می‌شود یا پس از نقطه‌ای مشخص، به دلیل همپوشانی اطلاعات و افزایش پیچیدگی، عملکرد سیستم وارد ناحیه اشباع می‌شود.

معیارهای ارزیابی عملکرد مدل‌ها

به‌منظور سنجش و مقایسه دقیق عملکرد مدل‌های منفرد و ترکیبی، از شاخص‌های آماری و معیارهای خطای زیر استفاده شد:

ریشه میانگین مربع خطا (RMSE)[1]: این معیار یکی از پرکاربردترین شاخص‌ها در ارزیابی دقت مدل‌های هیدرولوژیکی است که میزان پراکندگی یا انحراف مقادیر پیش‌بینی‌شده از مقادیر مشاهده‌شده را کمی‌سازی می‌کند و هرچه مقدار آن به صفر نزدیک‌تر باشد، بیانگر دقت و کارایی بالاتر مدل است.

رابطه ۶)
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - S_i)^2}{n}}$$

میانگین قدرمطلق خطا (MAE)[2]: این شاخص بیانگر میانگین قدرمطلق اختلاف بین مقدارهای مشاهده‌شده و مقدارهای پیش‌بینی‌شده است. مقدار کمتر MAE نشان‌دهنده دقت بالاتر مدل و نزدیکی بیشتر پیش‌بینی‌ها به مقادیر واقعی است.

رابطه ۷)
$$MAE = \frac{\sum_{i=1}^{n}(O_i - S_i)}{n}$$

ریشه میانگین مربعات خطا (MSE)[3]: این شاخص بیانگر میانگین مربع اختلاف میان مقادیر مشاهده‌شده و پیش‌بینی‌شده است. مقدار MSE همواره غیرمنفی بوده و هرچه به صفر نزدیک‌تر باشد، نشان‌دهنده دقت و برازش بالاتر مدل است.

رابطه ۸)
$$MSE = \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{n}$$

[1] RMSE: Root Mean Square Error
[2] MAE: Mean Absolute Error
[3] MSE: Mean Square Error

بایاس (Bias): شاخص بایاس میزان تمایل مدل به بیش‌برآوردی یا کم‌برآوردی مقادیر را نشان می‌دهد. مقدار مثبت بیانگر بیش‌برآوردی[1] و مقدار منفی نشان‌دهنده کم‌برآوردی[2] مدل است و هرچه به ۰ نزدیک‌تر باشد نزدیک تر بودن به مقادیر واقعی را نشان می‌دهد.

$$SE = \bar{O} - \bar{S}$$ رابطه ۹)

شاخص نش–ساتکلیف (NSE[3]): این شاخص یکی از معیارهای پرکاربرد در ارزیابی مدل‌های هیدرولوژیکی است و میزان کارایی مدل در بازتولید داده‌های مشاهده‌ای را نشان می‌دهد. مقدار این شاخص بین منفی بی‌نهایت تا ۱ تغییر می‌کند و هرچه مقدار آن به ۱ نزدیک‌تر باشد، نشان‌دهنده عملکرد بهتر مدل است.

$$NSE = \frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2}$$ رابطه ۱۰)

O_i مقدار مشاهده شده در گام زمانی i

$\bar{O}$ میانگین مقدار مشاهداتی

$\bar{S}$ میانگین مقادیر پیشبینی شده

S_i مقدار پیشبنی شدت در گام زمانی i

n تعداد نمونه یا تعداد داده

کلیه این معیارها برای تمام مدل‌های منفرد و ترکیبی محاسبه شد تا بهترین ساختار مدل‌سازی بارش ماهانه در محدوده مطالعه انتخاب گردد.

یافته‌های پژوهش

نتایج ارزیابی دقت ماهواره

به‌منظور صحت‌سنجی داده‌های ماهواره‌ای CHIRPS در محدوده حوضه آبریز دریاچه نمک، از داده‌های مشاهده‌ای چهار ایستگاه باران‌سنجی منتخب شامل ایستگاه‌های آغچه (استان اصفهان)، ارتش‌آباد (استان قزوین)، و ایستگاه‌های آراقلعه و احمدآباد (استان مرکزی) بهره‌گیری شد. این ایستگاه‌ها بر اساس دوره آماری ۱۰ ساله (۲۰۰۷ تا ۲۰۱۷) انتخاب و داده‌های بارش ماهانه آن‌ها استخراج گردید. در ادامه، عملکرد داده‌های CHIRPS از طریق مقایسه مستقیم با داده‌های زمینیِ متناظر در مقیاس زمانی ماهانه مورد ارزیابی قرار گرفت که خلاصه نتایج این تحلیل در جدول۳ گزارش شده است.

جدول ۳. نتایج ارزیابی دقت داده‌های CHIRPS در منطقه مورد مطالعه

منبع داده	همبستگی (-)	بایاس (mm)	میانگین توان دو خطا (٪)	میانگین قدر مطلق خطا (٪)	نش–سایتیف (٪)
CHIRPS	۷۱/۶۷	-۰۰/۰۹	۱۹/۴۲	۳۱/۳۶	۷۰/۰۰

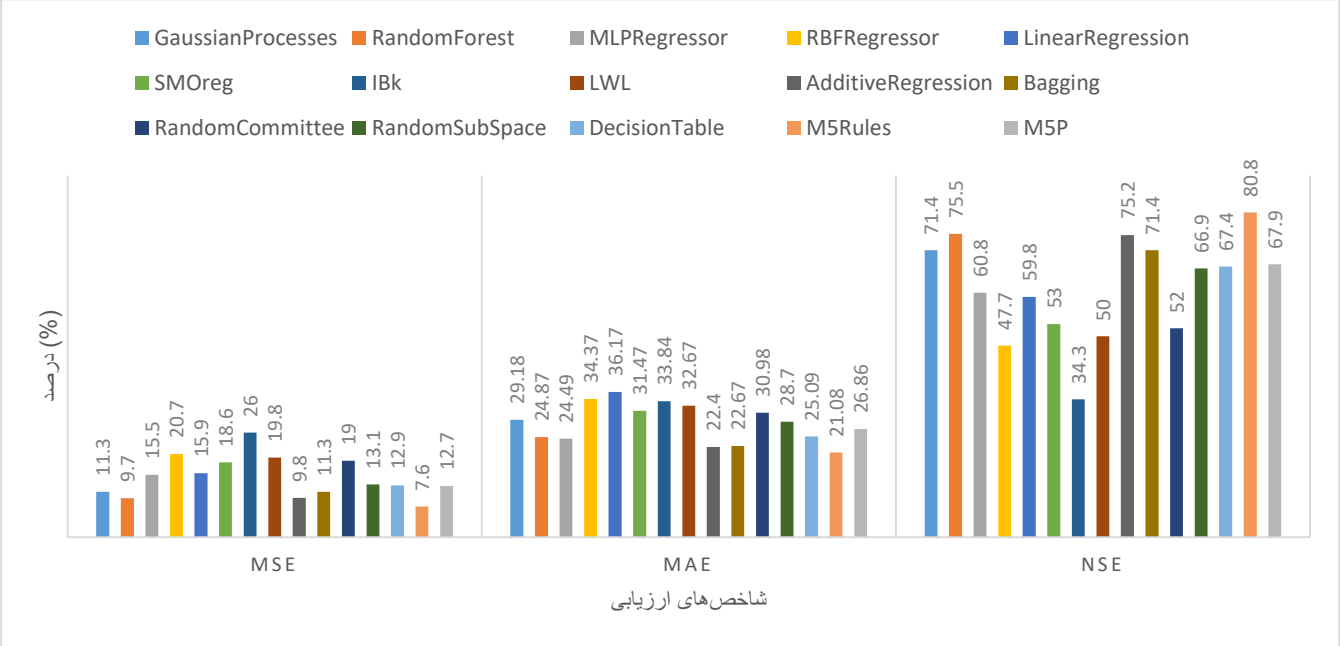
نتایج ارزیابی آماری، کارایی محصول بارش CHIRPS را در بازتولید الگوهای بارش ماهانه حوضه مورد مطالعه تأیید می‌کند. مقدار شاخص کارایی نش–ساتکلیف (NSE) معادل درصد۷۰، عملکرد قابل‌قبول و دقت قابل‌اتکای این مجموعه داده را در شبیه‌سازی رفتار بارش حوضه نشان می‌دهد. مقادیر خطای مطلق میانگین (MAE) و خطای میانگین مربعات (MSE) به‌ترتیب ۳۱/۳۶ و ۱۹/۴۲ درصد

[1] Overestimation
[2] Underestimation
[3] NSE: Nash–Sutcliffe Efficiency

محاسبه شدند که ضمن تأیید همخوانی مطلوب، بیانگر وجود درصدی از خطا در برآوردهاست. علاوه بر این، مقدار شاخص بایاس (Bias) برابر با ۰۰/۰۹- نشان‌دهنده گرایش ان CHIRPS به کم‌برآوردی بارش در مقایسه با داده‌های ایستگاهی است. با این وجود، با توجه به شرایط اقلیمی خشک و نیمه‌خشک حوضه و پراکندگی جغرافیایی ایستگاه‌های زمینی، دقت حاصل برای یک محصول ماهواره‌ای مطلوب ارزیابی شده و استفاده از داده‌های CHIRPS به عنوان مبنایی قابل‌اطمینان برای مدل‌سازی‌های هیدرولوژیک این حوضه کاملاً توجیه‌پذیر است.

عملکرد مدل های منفرد

در این مرحله، عملکرد ۱۵ الگوریتم یادگیری ماشین در پیش‌بینی بارش ماهانه حوضه آبریز دریاچه نمک به‌صورت منفرد مورد ارزیابی قرار گرفت. ارزیابی دقت مدل‌ها با بهره‌گیری از داده‌های دوره آزمون (ژانویه ۲۰۲۲ تا دسامبر ۲۰۲۴) و بر اساس شاخص‌های آماری متداول شامل RMSE، MAE، MSE و NSE انجام پذیرفت. نتایج کمی حاصل از فرآیند پیش‌بینی مدل‌های مذکور جهت مقایسه و تحلیل دقیق‌تر در شکل ۳ ارائه شده است.



شکل ۳. ارزیابی عملکرد مدل‌ها با شاخص‌های NSE, MAE و MSE

نتایج ارائه‌شده در شکل ۳ نشان می‌دهد که عملکرد مدل‌های یادگیری ماشین تنها به یک شاخص آماری وابسته نیست، بلکه حاصل برهم‌کنش هم‌زمان خطاهای تصادفی، خطاهای سیستماتیک و توانایی مدل در تعمیم الگوهای موجود در داده‌های بارش است. به همین دلیل، ارزیابی هم‌زمان شاخص‌های MSE، MAE، RMSE، Bias و NSE تصویر جامع‌تری از رفتار هر مدل ارائه می‌دهد و از نتیجه‌گیری بر اساس یک معیار منفرد جلوگیری می‌کند. در میان مدل‌های بررسی‌شده، M5Rules با مقدار NSE برابر با ۰/۸۰ درصد و کمترین مقدار MSE، بهترین عملکرد را در پیش‌بینی بارش ماهانه نشان داد. این موضوع بیانگر آن است که این مدل توانسته ضمن کاهش خطاهای بزرگ، الگوی تغییرات زمانی بارش را نیز با دقت مناسبی بازتولید کند. همچنین مدل AdditiveRegression نیز با وجود ساختار متفاوت، عملکردی نزدیک به M5Rules داشت که می‌تواند ناشی از توانایی آن در کاهش بایاس و بهبود تدریجی عملکرد مدل پایه باشد. در مقابل، مدل IBk ضعیف‌ترین عملکرد را از نظر NSE نشان داد. اگرچه مقدار MAE این مدل نسبت به

برخی مدل‌های دیگر تفاوت چشمگیری نداشت، اما افزایش قابل‌توجه MSE بیانگر آن است که این الگوریتم در برخی بازه‌های زمانی با خطاهای بزرگ مواجه شده است. از آنجا که MSE نسبت به خطاهای بزرگ حساسیت بیشتری دارد، این نتیجه نشان می‌دهد که مدل IBk در بازتولید رخدادهای شدید بارش از پایداری کافی برخوردار نبوده است.

مقایسه هم‌زمان MAE و MSE نیز اطلاعات ارزشمندی درباره ماهیت خطاهای مدل‌ها ارائه می‌دهد. نزدیک بودن مقادیر این دو شاخص در مدل‌هایی مانند M5Rules و AdditiveRegression نشان می‌دهد که توزیع خطاها نسبتاً یکنواخت بوده و مدل‌ها کمتر تحت تأثیر مقادیر پرت قرار گرفته‌اند. در مقابل، اختلاف بیشتر بین MAE و MSE در مدل‌هایی مانند IBk و RBFRegressor حاکی از وجود خطاهای بزرگ در بخشی از نمونه‌هاست که موجب افزایش MSE شده است. بنابراین، استفاده هم‌زمان از این دو شاخص امکان تشخیص پایداری مدل‌ها را با دقت بیشتری فراهم می‌کند.

بررسی شاخص Bias نیز نشان داد که مدل Gaussian Processes با مقدار بایاس نزدیک به ۰/۰۴ کمترین انحراف سیستماتیک را در برآورد مقادیر بارش داشته است؛ به این معنا که میانگین مقادیر پیش‌بینی‌شده این مدل بیشترین انطباق را با میانگین داده‌های مشاهده‌شده دارد. در مقابل، مدل Random SubSpace با مقدار بایاس حدود ۰/۱۹ بیشترین تمایل به بیش‌برآوردی یا کم‌برآوردی سیستماتیک را نشان داد. هرچند مقدار کم بایاس لزوماً به معنای بهترین عملکرد کلی نیست، اما بیانگر آن است که مدل از نظر برآورد میانگین بارش، رفتار متعادلی داشته است.

تفاوت عملکرد مدل‌ها را می‌توان با توجه به ویژگی‌های الگوریتمی آن‌ها نیز تبیین کرد. مدل‌های مبتنی بر درخت تصمیم، مانند Random Forest و M5Rules، به دلیل توانایی در مدل‌سازی روابط غیرخطی، مقاومت در برابر نویز و قابلیت استخراج قوانین تصمیم، توانسته‌اند تغییرات پیچیده بارش را با دقت بیشتری بازنمایی کنند. در مقابل، عملکرد نسبتاً ضعیف‌تر مدل MLPRegressor را می‌توان به محدود بودن حجم داده‌های آموزشی و حساسیت این مدل به تنظیم پارامترها نسبت داد. شبکه‌های عصبی معمولاً برای دستیابی به قابلیت تعمیم مناسب به داده‌های آموزشی بیشتری نیاز دارند و در مجموعه داده‌های نسبتاً کوچک ممکن است دچار بیش‌برازش یا ناپایداری شوند. همچنین مدل‌های خطی مانند Linear Regression ، به دلیل فرض رابطه خطی میان متغیرها، در بازنمایی رفتار غیرخطی بارش محدودیت دارند و در نتیجه عملکرد ضعیف‌تری نسبت به مدل‌های غیرخطی نشان داده‌اند.

در مجموع، نتایج نشان می‌دهد که برتری یک مدل تنها به کاهش مقدار یک شاخص خطا محدود نمی‌شود، بلکه حاصل تعادل میان کاهش خطاهای بزرگ، کنترل بایاس، توانایی مدل در استخراج روابط غیرخطی و قابلیت تعمیم آن به داده‌های آزمون است. این موضوع اهمیت استفاده از چندین شاخص ارزیابی و تحلیل هم‌زمان آن‌ها را در انتخاب مدل مناسب برای پیش‌بینی بارش ماهانه آشکار می‌سازد.

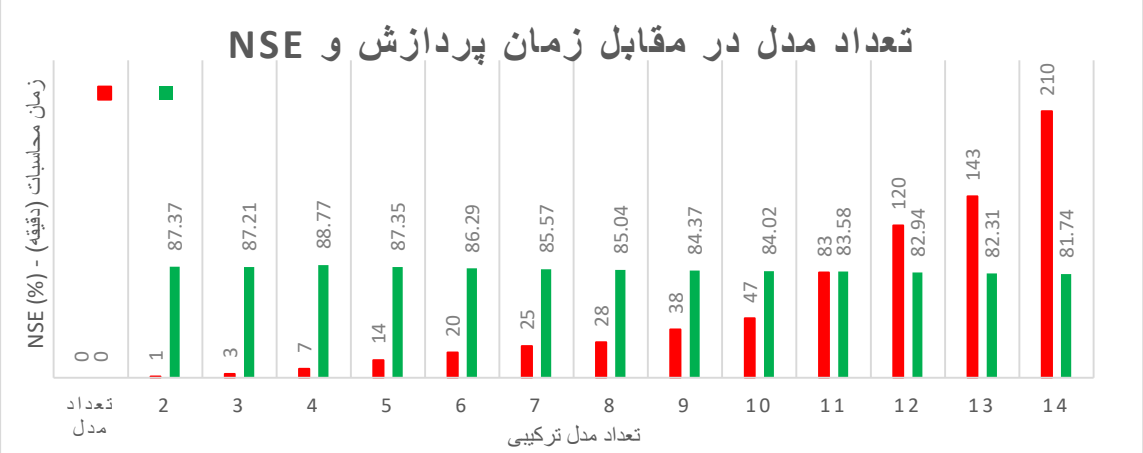
اثر ترکیب مدل‌ها

در فرآیند تحلیل تمامی ترکیبات ممکن از مدل‌های یادگیری ماشین، ابتدا هر یک از مدل‌های پایه به‌صورت مستقل آموزش داده شدند و مقادیر پیش‌بینی‌شده آن‌ها برای دوره آزمون استخراج گردید. سپس، برای هر ترکیب از مدل‌ها، خروجی مدل‌های منتخب با استفاده از روش انسامبل وزن‌دار تلفیق شد؛ به‌گونه‌ای که سهم هر مدل در پیش‌بینی نهایی متناسب با وزن بهینه آن تعیین گردید. وزن‌های هر ترکیب با استفاده از روش جستجوی فراگیر و بر اساس بیشینه‌سازی شاخص NSE در دوره آزمون تعیین شدند و پیش‌بینی نهایی از میانگین وزنی خروجی مدل‌های عضو انسامبل به دست آمد. در ادامه در جدول ۴ و برای نمایش آسان‌تر در شکل ۴ اثر تعداد مدل در زمان پردازش و دقت مدل بررسی می‌شود.

جدول ۴. اثر تعداد مدل ترکیبی در مقابل زمان پردازش و NSE

تعداد مدل	تعداد ترکیب	زمان محاسبات	ببیشترین NSE	کمترین RMSE

		(دقیقه)	(٪)	(٪)
۲	۱۰۵	۱≈	۸۷/۳۷	۲۲/۳۷
۳	۴۵۵	۳≈	۸۷/۲۱	۲۲/۵۱
۴	۱۳۶۵	۷≈	۸۸/۷۷	۲۱/۱۰
۵	۳۰۰۳	۱۴≈	۸۷/۳۵	۲۲/۳۹
۶	۵۰۰۵	۱۹≈	۸۶/۲۹	۲۳/۳۱
۷	۶۴۳۵	۲۳≈	۸۵/۵۷	۲۳/۹۱
۸	۶۴۳۵	۲۵≈	۸۵/۰۴	۲۴/۳۵
۹	۵۰۰۵	۳۵≈	۸۴/۳۷	۲۴/۸۸
۱۰	۳۰۰۳	۵۷≈	۸۴/۰۲	۲۵/۱۶
۱۱	۱۳۶۵	۸۰≈	۸۳/۵۸	۲۵/۵۱
۱۲	۴۵۵	۱۲۰≈	۸۲/۹۴	۲۶/۰۰
۱۳	۱۰۵	۱۳۶≈	۸۲/۳۱	۲۶/۴۸
۱۴	۱۵	۲۰۰≈	۸۱/۷۴	۲۶/۹۰
۱۵	۱	۳۲۰≈	۸۱/۱۸	۲۷/۳۱



شکل ۴. اثر تعداد مدل ترکیبی در مقابل زمان پردازش و NSE

بررسی نتایج جدول ۴ و شکل ۴ نشان می‌دهد که عملکرد مدل‌های انسامبل به‌طور قابل توجهی تحت تأثیر تعداد مدل‌های مشارکت‌کننده در ترکیب قرار دارد و این رابطه رفتاری غیرخطی از خود نشان می‌دهد. همان‌گونه که مشاهده می‌شود، با افزایش تعداد مدل‌ها از ۲ به ۴، دقت پیش‌بینی بهبود یافته و شاخص‌های ارزیابی عملکرد روند مطلوبی را نشان می‌دهند. در این میان، بهترین عملکرد در ترکیب چهار مدل حاصل شده است؛ به‌طوری که بیشترین مقدار شاخص NSE برابر با ۸۸/۷۷ درصد و کمترین مقدار RMSEمعادل ۲۱/۱۰ درصد به دست آمده است. این موضوع بیانگر آن است که ترکیب تعداد محدودی از مدل‌های کارآمد می‌تواند به شکل مؤثری نقاط ضعف یکدیگر را جبران کرده و خطای پیش‌بینی را کاهش دهد. با این حال، پس از این نقطه بهینه، افزایش تعداد مدل‌ها منجر به بهبود بیشتر عملکرد نشده و حتی روندی کاهشی در دقت پیش‌بینی مشاهده می‌شود. به‌عنوان نمونه، مقدار NSEاز ۸۷/۳۵ درصد در ترکیب پنج مدل به‌تدریج کاهش یافته و در ترکیب پانزده مدل به حدود ۸۱/۱۸ درصد می‌رسد، در حالی که RMSEنیز از مقدار بهینه ۲۱/۱۰ در چهار مدل به ۲۷/۳۱ درصد در ترکیب پانزده مدل افزایش می‌یابد. این روند نشان می‌دهد که ورود مدل‌های بیشتر به ترکیب لزوماً به معنای افزایش کیفیت پیش‌بینی نیست و در برخی موارد می‌تواند به دلیل افزایش نویز پیش‌بینی‌ها یا همپوشانی الگوهای خطا میان مدل‌ها موجب کاهش کارایی شود.

با وجود آنکه نتایج کمی این پژوهش بر اساس داده‌های یک منطقه مشخص به دست آمده است، انتظار نمی‌رود که مقادیر دقیق شاخص‌های عملکرد، تعداد بهینه مدل‌ها یا زمان محاسباتی بدون تغییر برای تمامی مناطق قابل تعمیم باشند؛ زیرا این پارامترها به ویژگی‌های اقلیمی، طول دوره آماری، کیفیت داده‌ها، الگوی زمانی بارش و میزان غیرخطی بودن روابط بین متغیرها وابسته هستند. با این حال، انتظار می‌رود الگوی کلی مشاهده‌شده در این پژوهش، مبنی بر وجود یک نقطه بهینه برای اندازه انسامبل و کاهش بازده ناشی از افزایش بیش از حد تعداد مدل‌ها، در بسیاری از مناطق با شرایط اقلیمی مشابه نیز برقرار باشد. دلیل این امر آن است که پدیده اشباع عملکرد در انسامبل‌ها ناشی از اصول بنیادین یادگیری ماشین، از جمله همبستگی میان مدل‌ها، کاهش تنوع مؤثر اعضای انسامبل و افزایش اطلاعات تکراری است و کمتر به ویژگی‌های یک منطقه خاص وابسته است. بنابراین، چارچوب پیشنهادی این پژوهش برای تعیین تعداد بهینه مدل‌های انسامبل می‌تواند در سایر حوضه‌های آبریز و مناطق دارای اقلیم مشابه نیز مورد استفاده قرار گیرد، هرچند تعیین تعداد بهینه مدل‌ها و مقادیر دقیق شاخص‌های عملکرد، مستلزم آموزش و ارزیابی مجدد مدل‌ها بر اساس داده‌های همان منطقه خواهد بود..

جدول ۵. توزیع فراوانی شاخص NSE برای تعداد ترکیب های مختلف

تعداد مدل ترکیب شده	بازه‌های NSE								
	۴۵–۵۰	۵۰–۵۵	۵۵–۶۰	۶۰–۶۵	۶۵–۷۰	۷۰–۷۵	۷۵–۸۰	۸۰–۸۵	۸۵–۹۰
۲	۲	۱	۱	۷	۲۲	۳۵	۲۲	۱۴	۱
۳	۱	۰	۴	۹	۳۹	۱۳۷	۱۷	۸۳	۱۲
۴	۰	۰	۲	۸	۴	۳۱	۶۳۴	۳۵۲	۲۸
۵	۰	۰	۰	۰	۴۱	۳۸۷	۱۵۵۶	۹۸۳	۳۴
۶	۰	۰	۰	۰	۲۳	۳۶	۲۶۱۵	۱۹۵۸	۴۹
۷	۰	۰	۰	۰	۳	۲۵۱	۳۱۸۵	۲۹۶۸	۲۸
۸	۰	۰	۰	۰	۰	۱۸	۲۹۲	۳۴۶	۱
۹	۰	۰	۰	۰	۰	۲۶	۲۶	۲۹۷۳	۰
۱۰	۰	۰	۰	۰	۰	۲	۱۱۴	۱۹۸۷	۰
۱۱	۰	۰	۰	۰	۰	۰	۳۶۹	۹۹۶	۰
۱۲	۰	۰	۰	۰	۰	۰	۹۳	۳۶۲	۰
۱۳	۰	۰	۰	۰	۰	۰	۱۳	۹۲	۰
۱۴	۰	۰	۰	۰	۰	۰	۲	۱۳	۰
۱۵	۰	۰	۰	۰	۰	۰	۰	۱	۰

این جدول توزیع فراوانی مقادیر شاخص NSE را برای ترکیب‌های مختلف مدل‌های یادگیری ماشین (از ۲ تا ۱۵ مدل) نشان می‌دهد و در حقیقت چشم‌انداز کلی دقت و پایداری کل فرآیند شبیه‌سازی را ترسیم می‌کند. الگوی مشاهده‌شده بیانگر آن است که با تغییر تعداد مدل‌های انسامبل، چگالی احتمال دستیابی به دقت‌های بالاتر چگونه جابه‌جا شده و سیستم چه مسیری از پراکندگی اولیه تا تمرکز و سپس اشباع را طی کرده است.

در ترکیب‌های کوچک‌تر (۲ تا ۴ مدل)، پراکندگی مقادیر NSE بسیار زیاد است و عملکرد مدل‌ها در طیفی گسترده، از بازه‌های ضعیف تا متوسط و حتی نسبتاً خوب توزیع شده است. با ورود به ترکیب‌های ۵ تا ۷ مدل، ساختار توزیع به‌شدت منظم‌تر می‌شود و مرکز ثقل آن به‌طور قابل توجهی به سمت بازه‌های ۷۵–۸۵ و حتی فراتر از آن حرکت می‌کند. این جابه‌جایی به روشنی نشان می‌دهد که افزایش تعداد مدل‌ها در این محدوده، موجب تقویت پایداری پیش‌بینی‌ها و افزایش احتمال دستیابی به دقت‌های بالا می‌شود. اما این

روند افزایشی تنها تا یک نقطه ادامه دارد. از ترکیب حدود ۷ مدل به بالا، سیستم وارد مرحله‌ای شبیه به تغییر فاز می‌شود؛ جایی که اگرچه همچنان برخی ترکیب‌ها عملکرد قابل‌قبولی دارند، اما تعداد آن‌ها به‌طور محسوس کاهش می‌یابد. در ترکیب‌های ۹ مدل به بالا، هم‌زمان با کاهش چشمگیر تعداد ترکیب‌های ممکن، توان دستیابی به NSE های بالا (به‌ویژه در بازه ۸۵–۹۰) تقریباً از بین می‌رود. این مرحله نوعی "اشباع" عملکردی است که در آن مدل‌ها به جای تقویت یکدیگر، دچار هم‌پوشانی اطلاعات و تمایل به میانگین‌گیری افراطی می‌شوند؛ پدیده‌ای که میتوان آنرا تداخل اطلاعات نامید. نتیجه این است که انسامبل‌های بسیار بزرگ عمدتاً در بازه‌های دقت میانی گیر می‌کنند و قابلیت رسیدن به مقادیر استثنایی را که در ترکیب‌های کوچک‌تر مشاهده می‌شد، از دست می‌دهند.

این تحلیلِ توزیع فراوانی به‌وضوح نشان می‌دهد که چرا ترکیب ۴ مدل به‌عنوان گزینه بهینه انتخاب شده است. این ترکیب نه تنها رکوردهای عملکردی خوبی ثبت کرده، بلکه از نظر احتمال دستیابی به دقت‌های مطلوب در بازه‌های ۷۵–۹۰ نیز توزیع متوازن و کارآمدتری نسبت به ترکیب‌های بزرگ‌تر دارد. به تعبیر دیگر، نقطه بهینه نه جایی است که بیشترین تعداد مدل‌ها حضور دارند، بلکه جایی است که بهترین توازن میان کیفیت غالب، احتمال دستیابی به دقت بالا و اجتناب از اشباع مدل برقرار شده است.

بحث

نتایج این پژوهش نشان داد که رابطه میان تعداد اعضای انسامبل و دقت پیش‌بینی بارش خطی و افزایشی نیست و پس از رسیدن به اندازه‌ای مشخص، عملکرد به حالت اشباع نزدیک می‌شود. انسامبل چهارمدلی با NSE برابر با ۸۸,۷۶٪ و RMSE معادل ۲۱,۱۰، بهترین تعادل میان دقت پیش‌بینی، پایداری مدل و هزینه محاسباتی را ایجاد کرد. افزایش تعداد مدل‌ها فراتر از این نقطه، به دلیل ورود اطلاعات تکراری و محدود شدن سهم اعضای جدید، بهبود قابل‌توجهی در عملکرد ایجاد نکرد. این یافته با نتایج Na et al. (2019)، Ho et al. (2021) و Palma et al. (2025) همخوان است که نشان دادند عملکرد انسامبل‌ها ابتدا با افزایش اعضا بهبود یافته، اما پس از یک نقطه بهینه، بازده نهایی کاهش می‌یابد و هزینه پردازشی افزایش پیدا می‌کند. یافته‌های پژوهش همچنین نشان داد که تنوع ساختاری مدل‌ها اهمیت بیشتری از افزایش صرف تعداد اعضا دارد. عملکرد مطلوب انسامبل چهارمدلی را می‌توان به توانایی مدل‌های منتخب در پوشش الگوهای متفاوت خطا و ارائه اطلاعات مکمل نسبت داد. این نتیجه با Sharma et al. (2019) و Darbandsari et al. (2019) مطابقت دارد که افزایش مهارت انسامبل را بیشتر ناشی از تفاوت و مکمل بودن مدل‌ها می‌دانند. Velazquez et al. (2010) نیز نشان دادند که انسامبل بهینه الزاماً از بهترین مدل‌های منفرد تشکیل نمی‌شود و ترکیب مدل‌های مکمل می‌تواند عملکرد بهتری ایجاد کند. در روش وزن‌دهی مورد استفاده، اختصاص وزن‌های بهینه به مدل‌های پایه امکان بهره‌گیری از اطلاعات مکمل و کاهش اثر خطاهای اختصاصی را فراهم می‌کند. با افزایش تعداد اعضا بدون افزایش تنوع، همبستگی خطاها و افزونگی اطلاعات افزایش یافته و اعضای جدید سهم محدودی در کاهش خطا خواهند داشت. در نتیجه، پیچیدگی، زمان آموزش و هزینه محاسباتی افزایش می‌یابد. این رفتار با مفهوم مرز بهینه پارتو سازگار است؛ یعنی سامانه مطلوب الزاماً سامانه‌ای با بیشترین تعداد مدل نیست، بلکه سامانه‌ای است که با مدل‌های متنوع و مکمل، بیشترین دقت را با کمترین پیچیدگی و هزینه فراهم کند.

نتایج همچنین ظرفیت مناسب داده‌های ماهواره‌ای CHIRPS را برای توسعه سامانه‌های پیش‌بینی بارش مبتنی بر یادگیری ماشین و انسامبل نشان داد. این ویژگی به‌ویژه در مناطقی با کمبود ایستگاه‌های زمینی یا داده‌های ناقص اهمیت دارد. در مجموع، یافته‌ها نشان می‌دهند که موفقیت سامانه‌های انسامبل بیش از اندازه آن‌ها به کیفیت طراحی، تنوع اعضا و انتخاب اندازه بهینه وابسته است؛ بنابراین، افزایش تعداد مدل‌ها بدون توجه به تنوع ساختاری، لزوماً به افزایش دقت منجر نمی‌شود.

نتیجه‌گیری

هدف اصلی این پژوهش، بررسی اثر تعداد مدل‌های یادگیری ماشین بر عملکرد ترکیب و انسامبل‌سازی در پیش‌بینی بارش و تعیین تعداد بهینه مدل‌ها بود. نتایج نشان داد که رابطه میان تعداد اعضای انسامبل و دقت پیش‌بینی غیرخطی است و افزایش تعداد مدل‌ها لزوماً به بهبود عملکرد منجر نمی‌شود. داده‌های CHIRPS نیز عملکرد قابل قبولی در بازنمایی بارش نشان دادند.

با افزایش تعداد مدل‌ها از ۲ به ۴، عملکرد پیش‌بینی بهبود یافت و ترکیب چهارمدلی شامل SMOreg ، AdditiveRegression، DecisionTableو M5Rules بهترین عملکرد را با NSE برابر ۰/۸۸ و RMSE برابر ۲۱/۱۰ به دست آورد. پس از این نقطه، افزایش تعداد اعضا موجب اشباع عملکرد، هم‌پوشانی اطلاعات و افزایش خطا شد؛ به‌گونه‌ای که در ترکیب ۱۵ مدلی، NSEبه حدود ۰،۸۱ کاهش یافت، در حالی که هزینه محاسباتی از حدود ۱ دقیقه به ۳۲۰ دقیقه افزایش پیدا کرد. بنابراین، ترکیب چهارمدلی نقطه بهینه‌ای میان دقت، پایداری و هزینه محاسباتی ایجاد کرد. این یافته‌ها نشان می‌دهند که فرض «افزایش تعداد مدل‌ها برابر با افزایش دقت» همواره معتبر نیست و تنوع و مکمل‌بودن مدل‌ها می‌تواند از تعداد صرف اعضای انسامبل مهم‌تر باشد. از این منظر، انتخاب تعداد محدودی از مدل‌های کارآمد می‌تواند در سامانه‌های عملیاتی مدیریت منابع آب و پیش‌بینی رویدادهایی مانند سیلاب، دقت مناسب را با هزینه محاسباتی کمتر فراهم کند. مهم‌ترین محدودیت پژوهش، وابستگی به داده ماهواره‌ای CHIRPS و احتمال وجود خطاهای سیستماتیک در مناطق کم‌برخوردار از داده‌های زمینی است. همچنین، اگرچه تعداد مدل‌ها، دقت و هزینه محاسباتی بررسی شد، اثر تنوع ساختاری مدل‌ها و همبستگی خطاهای اعضای انسامبل به‌صورت کمی تحلیل نشد.

در مطالعات آینده، پیشنهاد می‌شود پایداری اندازه بهینه انسامبل در حوضه‌های با شرایط اقلیمی و توپوگرافی متفاوت و در دوره‌های خشکسالی و ترسالی شدید ارزیابی شود. همچنین، استفاده از وزن‌دهی پویا و تطبیقی، بررسی محصولات مختلف ماهواره‌ای و داده‌های تلفیقی ماهواره–ایستگاه، و به‌کارگیری الگوریتم‌های یادگیری عمیق و روش‌های پیشرفته انسامبل می‌تواند به بهبود عملکرد و قابلیت تعمیم نتایج کمک کند. بررسی کمی تنوع مدل‌ها، همبستگی خطاها و نقش آن‌ها در ایجاد اشباع عملکرد نیز می‌تواند سازوکار تعیین اندازه بهینه انسامبل را روشن‌تر سازد.

در مجموع، پیام راهبردی پژوهش آن است که در طراحی سامانه‌های پیش‌بینی، بهینگی بر کثرت اولویت دارد؛ بنابراین، تعیین تعداد مناسب اعضای انسامبل یک تصمیم مهندسی مهم است که می‌تواند هم‌زمان دقت و پایداری پیش‌بینی را افزایش داده و از تحمیل هزینه‌های محاسباتی غیرضروری جلوگیری کند.

مراجع

دلیر گبرآباد، علی، عبداللهی، محمدامین، اشرف‌زاده، افشین. (۱۴۰۵). شبیه‌سازی بارندگی ماهانه با استفاده از مدل‌های یادگیری ماشین ترکیبی: حوضه دریاچه نمک. نشریه آب و خاک، ۸(۳)، ۴۷۷-۴۹۸. https://doi.org/10.22067/jsw.2026.97595.1522

عبداللهی، محمدامین، عابدی‌کوپایی جهانگیر، متین‌زاده، محمدمهدی . (۱۴۰۳). تأثیر مساحت زیرحوضه‌ها و روش‌های محاسبه زمان تمرکز بر شبیه‌سازی حجم رواناب شهری با استفاده از نرم‌افزار SewerGEMS مطالعه موردی: شهرکرد. مجله علوم آب و خاک، ۳(۲۸)، ۱۵-۱. https://doi.org/10.47176/jwss.28.3.49834

REFERENCES

Abdollahi, M., Abedi, K. J., & Matinzadeh, M. (2024). The Effect of the Sub-Basins Area and Methods of Calculating Concentration Time on the Simulation of Urban Runoff Volume Using SewerGEMS Software.(Case Study: Shahrekord). *Journal of Water and Soil Science*. https://doi.org/10.47176/jwss.28.3.49834. (inPersian).

A. Zhai, G. Fan, Xiaowen Ding, G. Huang (2022). Regression Tree Ensemble Rainfall–Runoff Forecasting Model and Its Application to Xiangxi River, China. Water. https://doi.org/10.3390/w14030463.

Alpaydin, E. (2020). Introduction to Machine Learning (4th ed.). *MIT Press*.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., Huffman, G. J., Adler, R. F., & Wood, E. F. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, 23(1), 207–224. https://doi.org/10.5194/hess-23-207-2019, 2019.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197-227. https://doi.org/10.1007/s11749-016-0481-7.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). https://doi.org/10.1145/2939672.2939785.

C. M. DeChant, H. Moradkhani (2011). Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-15-3399-2011.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215. https://doi.org/10.1016/j.neucom.2019.10.118.

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14(2), 241-258. https://doi.org/10.1007/s11704-019-8208-z.

Dalir Gabrabad, A., Abdollahi, M. A., & Ashrafzadeh, A. (2026). Monthly rainfall simulation using hybrid machine learning models: Salt Lake Basin. *Journal of Water and Soil*. https://doi.org/10.22067/jsw.2026.97595.1522 (inPersian)

Emixi Valdez, F. Anctil, M. Ramos (2021). Choosing between post-processing precipitation forecasts or chaining several uncertainty quantification tools in hydrological forecasting systems. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-26-197-2022, 2022.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... & Michaelsen, J. (2015). The Climate Hazards Infrared Precipitation with Stations (CHIRPS). *Scientific Data*, 2, 150066. https://doi.org/10.1038/sdata.2015.66.

Fikileni, S., & Wolski, P. (2022). Framework for implementation of the Pitman-WR2012 model in seasonal hydrological forecasting: a case study of Kraai River, South Africa. *Water SA*, 48(1), 62–74–62–74. https://doi.org/10.17159/wsa/2022.v48.i1.3891.

Fürnkranz, J., Gamberger, D., & Lavrač, N. (2012). Foundations of Rule Learning. Springer.

Georgia A. Papacharalampous, Hristos Tyralis (2022). A review of machine learning concepts and methods for addressing challenges in probabilistic hydrological post-processing and forecasting. *Frontiers in Water*. https://doi.org/10.3389/frwa.2022.961954.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. *MIT Press*. https://doi. org/10.4258/hir. 2016.22. 4.351.

Hao Shi, Ting-Ji Huang, Lu Han, De-Chuan Zhan, Han-Jia Ye (2025). One-Embedding-Fits-All: Efficient Zero-Shot Time Series Forecasting by a Model Zoo. *arXiv.org* . https://doi.org/10.48550/arXiv.2509.04208.

H. Cloke, F. Pappenberger, Paul Smith, F. Wetterhall (2017). How do I know if I've improved my continental scale flood early warning system?. *Environmental Research Letters*. https://doi .org/ 10.1088/1748-9326/aa625a.

H. McMillan, E. Hreinsson, M. Clark, S. Singh, C. Zammit, M. Uddstrom (2013). Operational hydrological data assimilation with the recursive ensemble Kalman filter. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-17-21-2013.

Holmes, G., Hall, M., & Frank, E. (1999). Generating rule sets from model trees. Australian Joint Conference on Artificial Intelligence (pp. 1-12). Springer. *Australasian joint conference on artificial intelligence*. http://dx.doi.org/10.1007/978-3-540-46695-6_55.

Hartanto, M.I. (2019). Integrating Multiple Sources of Information for Improving Hydrological Modelling: An Ensemble Approach.

H. Ho, Y. Chiang, Chengyi Lin, Hong-Yuan Lee, Cheng-Chia Huang (2021). Development of an Interdisciplinary Prediction System Combining Sediment Transport Simulation and Ensemble Method. *Water*. https://doi.org/10.3390/w13182588.

He, X., J. Luo, P. Li, G. Zuo and J. Xie (2020). A hybrid model based on variational mode decomposition and gradient boosting regression tree for monthly runoff forecasting. *Water Resources Management* 34(2): 865–884. https://doi.org/10.1007/s11269-020-02483-x.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning: with Applications in R (2nd ed.). Springer. *academia.edu.* https://doi.org/10.1007/978-1-4614-7138-7.

J. Rasmussen, H. Madsen, K. Jensen, J. Refsgaard (2015). Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: evaluating the effect of ensemble size and localization on filter performance. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-19-2999-2015.

J. Frame, R. Araki, Soelem Aafnan Bhuiyan, Tadd Bindas, Jeremy Rapp, Lauren Bolotin et al (2025). Machine Learning for a Heterogeneous Water Modeling Framework. *JAWRA Journal of the American Water Resources Association*. https://doi.org/10.1111/1752-1688.70000.

Joseph Bellier, G. Bontron, I. Zin (2021). Selecting components in a probabilistic hydrological forecasting chain: the benefits of an integrated evaluation. *LHB*. https://doi.org/10.1080/27678490.2021.1936825.

Kundu, S., S. K. Biswas, D. Tripathi, R. Karmakar, S. Majumdar and S. Mandal (2023). A review on rainfall forecasting using ensemble learning techniques. *e-Prime-Advances in Electrical Engineering, Electronics and Energy* 6: 100296. https://doi.org/10.1016/j.prime.2023.100296.

Kuncheva, L. I. (2014). Combining Pattern Classifiers: Methods and Algorithms (2nd ed.). John Wiley & Sons.

Liyao Peng, Jiemin Fu, Yanbin Yuan, Xiang Wang, Yang Zhao, Jian Tong (2025). A Bayesian Ensemble Learning-Based Scheme for Real-Time Error Correction of Flood Forecasting. *Water*. https://doi.org/10.3390/w17142048.

Md Rasel Sheikh, Paulin Coulibaly (2024). Review of Recent Developments in Hydrologic Forecast Merging Techniques. *Water*. https://doi.org/10.3390/w16020301.

Patel, A., Keriwala, N., Soni, N., Goel, U., Bhoj, R., Adhyaru, Y., & Yadav, S. (2023). Rainfall Prediction using Machine Learning Techniques for Sabarmati River Basin, Gujarat, India. *Journal of Engineering Science & Technology* Review, 16(1). https://doi.org/ 10.25103/jestr.161.13

Palma, L., Peraza, A., Civantos, D., Duarte, A., Materia, S., Muñoz A.G., et al (2025). Data-driven seasonal climate predictions via variational inference and transformers. *npj Climate and Atmospheric Science*. https://doi.org/10.1038/s41612-026-01320-z.

Ramani, K., Reddy, M. S., Bhavani, K., Feeza, S., & Bavesh, V. S. (2024). Optimization of Rainfall Prediction Using Satellite Data Through Machine Learning and Deep Learning Algorithms. 2024 *IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems* (ICITEICS). https://doi.org/10.1109/ICITEICS61368.2024.10625624.

R. Figueiredo, K. Schröter, Alexander Weiss-Motz, M. Martina, H. Kreibich (2017). Improving accuracy and quantifying uncertainty in flood loss estimations through the use of multi-model ensembles. *Nat. Hazards Earth Syst. Sci. Discuss*. https://doi.org/10.5194/nhess-2017-349.

S. Baran, S. Hemri, Mehrez El Ayari (2018). Statistical post-processing of hydrological forecasts using Bayesian model averaging. *Water Resources Research*. https://doi.org/10.1029/2018WR024028.

Sikorska-Senoner, A.E., & Quilty, J. (2021). A novel ensemble-based conceptual-data-driven approach for improved streamflow simulations. *Environmental Modelling & Software*. https://doi.org/10.1016/j.envsoft.2021.105094.

Sharma, S., G. Raj Ghimire and R. Siddique (2023). Machine learning for postprocessing ensemble streamflow forecasts. *Journal of Hydroinformatics* 25(1): 126–139. https://doi.org/10.2166/hydro.2022.114.

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85, 1-16. https://doi.org/10.1016/j.jmp.2018.03.001.

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1249. https://doi.org/10.1002/widm.1249.

Snieder E., Karen Abogadil, U. Khan (2020). Resampling and ensemble techniques for improving ANN-based high streamflow forecast accuracy. *Hydrology and earth system sciences*. https://doi.org/10.5194/hess-25-2543-2021.

Sharma, S., Ridwan Siddique, S. Reed, P. Ahnert, A. Mejia (2019). Hydrological Model Diversity Enhances Streamflow Forecast Skill at Short- to Medium-Range Timescales. *Water Resources Research*. https://doi.org/10.1029/2018WR023197

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85, 1-16. https://doi.org/10.1016/j.jmp.2018.03.001.

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2). Springer. https://doi.org/10.1007/978-3-031-20719-8_2.

Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A brief review of nearest neighbor algorithm for learning and classification. *In 2019 International Conference on Intelligent Computing and Control Systems (ICCS)* (pp. 1255-1260). IEEE. https://doi.org/10.1109/ICCS45141.2019.9065747.

Tobias Necker, Ludwig Wolfgruber, Lukas Kugler, Martin Weissmann, Manfred Dorninger, S. Serafin (2024). The fractions skill score for ensemble forecast verification. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.4824.

Wu, Y., Wang, H., Zhang, B., & Du, K. L. (2012). Using radial basis function networks for function approximation and classification. *ISRN Applied Mathematics*. https://doi.org/10.5402/2012/324194.

Wooyoung Na, C. Yoo (2010). Optimize Short-Term Rainfall Forecast with Combination of Ensemble Precipitation Nowcasts by Lagrangian Extrapolation. *Water*.. https://doi.org/10.3390/w11091752.

William Armstrong, R. Arsenault, J. Martel, M. Troin, Patrice Dion, Behmard Sabzipour et al (2025). Improving multi-model ensemble streamflow forecasts by combining lumped, distributed and deep learning hydrological models. *Hydrological Sciences Journal*. https://doi.org/10.1080/02626667.2025.2471430.

Wonduk Seo, Juhyeon Lee, Yi Bu (2025). SPIO: Ensemble and Selective Strategies via LLM-Based Multi-Agent Planning in Automated Data Science. *arXiv.org*. https://doi.org/10.18653/v1/2026.acl-long.1039.

Wang, Y., & Witten, I. H. (1997). Induction of model trees for predicting continuous classes. *Proceedings of the poster papers of the European Conference on Machine Learning* (pp. 48-53). https://hdl.handle.net/10289/1183.